High Dimensional Kullback-Leibler divergence for grassland classification using satellite image time series with high spatial resolution

Maïlys Lopes$^1$, Stéphane Girard$^2$, Mathieu Fauvel$^1$

$^1$UMR Dynafor, INRA-INP Toulouse  $^2$Team Mistis, INRIA Grenoble

Journées de Statistique, Montpellier, 30 mai - 3 juin 2016
Outline

Introduction

High Dimensional Kullback-Leibler Divergence

Experimental results

Conclusion
Introduction

High Dimensional Kullback-Leibler Divergence

Experimental results

Conclusion
Objectives of the study

1. Characterization of grasslands through satellite image time series (SITS)
2. Classification of grassland management practices
3. Development of a statistical modelling of the grasslands signal suitable to SITS
Remote sensing imagery

A digital remote sensing image corresponds to a spatial, spectral and temporal sampling of a landscape.
Remote sensing imagery
A digital remote sensing image corresponds to a spatial, spectral and temporal sampling of a landscape.
Remote sensing imagery

A digital remote sensing image corresponds to a spatial, spectral and temporal sampling of a landscape.
Remote sensing imagery

A digital remote sensing image corresponds to a spatial, spectral and temporal sampling of a landscape.
Satellite remote sensing of grasslands

Grasslands are:

- Relatively **small** ($\approx 100m \times 100m$) ⇒ need **high spatial resolution**
- Heterogeneous ⇒ need **spectral bands**
- Have a **natural vegetation cycle disturbed by anthropic events** (mowing & grazing) ⇒ need **high temporal resolution**

⇒ Need dense **time series with high spatial resolution** to detect and to characterize grasslands.
Satellite image time series

Formosat-2

February

May

August

December
Normalized Difference Vegetation Index (NDVI)

**NDVI**: vegetation index that reflects the photosynthetic activity of the vegetation.

\[
\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}}, \quad -1 \leq \text{NDVI} \leq 1
\]
Representation of the grassland

A grassland is represented by a matrix of size \((n_i \times d)\), \(n_i\) varying with \(g_i\).

\[ x_{ik}(t) = \begin{pmatrix} x_{ik}(t_1) \\ \vdots \\ x_{ik}(t_d) \end{pmatrix} \in \mathbb{R}^d \]

with

- \(i\) grassland index,
- \(n_i\) number of pixels in grassland \(g_i\)
- \(k\) pixel index, \(k \in \{1, \ldots, n_i\}\)
- \(d\) length of time series
Statistical issues

A statistical model is required for:

- Processing grassland supervised classification
- Working at the parcel scale: Object-oriented classification
- Modelling grasslands with their constraints:
  - heterogeneity (spectral variability)
  - different size $n_i$
  - described by $d$ temporal variables
Introduction

High Dimensional Kullback-Leibler Divergence

Experimental results

Conclusion
Statistical modelling of grasslands

In each grassland $g_i$: each pixel $x_{ik} \sim \mathcal{N}(\mu_i, \Sigma_i)$

Maïlys Lopes
UMR Dynafor, INRA-INP Toulouse

High Dimensional Kullback-Leibler divergence for grassland classification using satellite image time series with high spatial resolution
Measuring proximity between two grasslands

Symmetrized Kullback-Leibler divergence:

\[
KLD(g_i, g_j) = \frac{1}{2} \left[ \text{Tr} \left( \Sigma_i^{-1} \Sigma_j + \Sigma_j^{-1} \Sigma_i \right) + (\mu_i - \mu_j)^\top (\Sigma_i^{-1} + \Sigma_j^{-1})(\mu_i - \mu_j) \right] - d
\]

- \( \Sigma_i \) is the covariance matrix,
- \( \mu_i \) is the mean vector of the signal,
- \( d \) is the number of variables,
- \( \text{Tr} \) is the trace operator.
Figure: Histogram of grassland size in number of pixels $n_i$. The red line corresponds to the number of parameters to estimate for each grassland for a multivariate Gaussian model. It is derived from the number of variables using the formula $d(d + 3)/2 = 170$ for $d = 17$. 

Maïlys Lopes  
UMR Dynafor, INRA-INP Toulouse
High dimensional model

According to High Dimensional Discriminant Analysis\(^1\) that assumes that the last eigenvalues of the covariance matrix are equal:

\[
\Lambda_i = \begin{pmatrix}
\lambda_{i1} & 0 & \cdots & 0 \\
0 & \ddots & \ddots & \vdots \\
0 & \ddots & \ddots & 0 \\
0 & \cdots & 0 & \lambda_{ip_i}
\end{pmatrix}
\]

where \(\lambda_{ij} \geq \lambda_i\), for \(j = 1, \ldots, p_i\)

---

High dimensional model

According to High Dimensional Discriminant Analysis\(^1\) that assumes that the last eigenvalues of the covariance matrix are equal:

\[
\Sigma_i = Q_i \Lambda_i Q_i^\top + \lambda_i I_d
\]

- \(I_d\) is the identity matrix of size \(d\),
- \(p_i\) is the number of non-equal eigenvalues,
- \(\lambda_i\) is the multiple eigenvalue corresponding to the noise term (last and equal eigenvalues),
- \(Q_i = [q_{i1}, \ldots, q_{ip_i}]\),
- \(\Lambda_i = \text{diag}\left[\lambda_{i1} - \lambda_i, \ldots, \lambda_{ip_i} - \lambda_i\right]\), \(q_{ij}, \lambda_{ij}\) are the \(j^{th}\) eigenvalues/eigenvectors of the covariance matrix \(\Sigma_i\), \(j \in \{1, \ldots, d\}\) such as \(\lambda_{i1} \geq \ldots \geq \lambda_{id}\).

High Dimensional Symmetrized KLD

Following this model, $\Sigma_i^{-1}$ can be computed explicitly:

$$
\Sigma_i^{-1} = -Q_i V_i Q_i^\top + \lambda_i^{-1} I_d
$$

with $V_i = \text{diag}\left[\frac{1}{\lambda_i} - \frac{1}{\lambda_{i1}}, \ldots, \frac{1}{\lambda_i} - \frac{1}{\lambda_{ip_i}}\right]$

Then:

$$
\text{HDKLD}(g_i, g_j) = \frac{1}{2} \left[ -\| \Lambda_j^{\frac{1}{2}} Q_j^\top Q_i V_i^{\frac{1}{2}} \|_F^2 - \| \Lambda_i^{\frac{1}{2}} Q_i^\top Q_j V_j^{\frac{1}{2}} \|_F^2 \\
+ \lambda_i^{-1} \text{Tr}\left[ \Lambda_j \right] - \lambda_j \text{Tr}\left[ V_i \right] + \lambda_j^{-1} \text{Tr}\left[ \Lambda_i \right] - \lambda_i \text{Tr}\left[ V_j \right] \\
- \| V_i^{\frac{1}{2}} Q_i^\top (\mu_i - \mu_j) \|^2 - \| V_j^{\frac{1}{2}} Q_j^\top (\mu_i - \mu_j) \|^2 \\
+ \frac{\lambda_i + \lambda_j}{\lambda_i \lambda_j} \|(\mu_i - \mu_j)\|^2 + \frac{\lambda_i^2 + \lambda_j^2}{\lambda_i \lambda_j} d \right] - d
$$

where $\| L \|^2_F = \text{Tr}(L^\top L)$ is the Frobenius norm.
Estimation

- $\hat{\lambda}_{ij}$ and $\hat{q}_{ij}$ are the first eigenvalues/eigenvectors of $\hat{\Sigma}_i$, $j \in \{1, \ldots, p_i\}$,

- $\hat{p}_i$ corresponds to the number of eigenvalues needed to reach a given percentage of variance $t$, \( \sum_{j=1}^{\hat{p}_i} \hat{\lambda}_{ij} \geq t \), $t$ being a user defined parameter,

- $\hat{\lambda}_i = \frac{\text{Tr}(\hat{\Sigma}_i) - \sum_{j \leq \hat{p}_i} \hat{\lambda}_{ij}}{d - \hat{p}_i}$.

Thus, to compute HDKLD only the $p_i$ first eigenvalues/eigenvectors are required and the unstable estimation of the eigenvectors associated to small eigenvalues is avoided.
Construction of a positive definite kernel

- (HD)KLD is a semi-metric.
- It can be turned to a positive definite kernel function.
- \[ K(g_i, g_j) = \exp \left[ - \frac{(HD)KLD(g_i, g_j)^2}{\sigma} \right] \text{ with } \sigma \in \mathbb{R}_+^* \]
Introduction

High Dimensional Kullback-Leibler Divergence

Experimental results

Conclusion
Data

**Study site**  Semi-rural area near Toulouse, France.

**Field data**  52 parcels with 3 management practices (field survey, 2015):

<table>
<thead>
<tr>
<th>Class</th>
<th>Nb of grasslands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mowing</td>
<td>34</td>
</tr>
<tr>
<td>Grazing</td>
<td>10</td>
</tr>
<tr>
<td>Mixed (mowing &amp; grazing)</td>
<td>8</td>
</tr>
<tr>
<td><strong>Total number of pixels:</strong></td>
<td><strong>8741</strong></td>
</tr>
</tbody>
</table>

**Satellite data**  Formosat-2

- Spatial resolution: 8m
- 4 spectral bands (B, G, R, NIR)
- Temporal frequency: up to 1 day
- \( d = 17 \) images from Feb. 16 to Dec. 20, 2013:
Classification methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Scale</th>
<th>Expl. variable</th>
<th>Kernel</th>
<th>Nb of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-SVM</td>
<td>Pixel</td>
<td>$x_{ik}$</td>
<td>RBF</td>
<td>8741</td>
</tr>
<tr>
<td>$\mu$-SVM</td>
<td>Object</td>
<td>$\mu_i$</td>
<td>RBF</td>
<td>52</td>
</tr>
<tr>
<td>KLD-SVM</td>
<td>Object</td>
<td>$\mathcal{N}(\mu_i, \Sigma_i)$</td>
<td>$K(g_i, g_j)$</td>
<td>52</td>
</tr>
<tr>
<td>HDKLD-SVM</td>
<td>Object</td>
<td>$\mathcal{N}(\mu_i, \Sigma_i)$</td>
<td>$K(g_i, g_j)$</td>
<td>52</td>
</tr>
</tbody>
</table>

+ majority rule

Optimal parameters have been optimized during cross-validation.
Processing chain

17 Formosat-2 images → NDVI computation → Smoothing (Whittaker Filter) [1] → Leave-One Out SVM classification (52 repetitions) → Training → Predicting one grassland label → 1 grassland out → Accuracy assessment

Field data
Results

Classification accuracy

<table>
<thead>
<tr>
<th></th>
<th>P-SVM</th>
<th>µ-SVM</th>
<th>KLD-SVM</th>
<th>HDKLD-SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRED</td>
<td>REF</td>
<td>REF</td>
<td>REF</td>
<td>REF</td>
</tr>
<tr>
<td></td>
<td>32 4 2</td>
<td>31 6 3</td>
<td>32 8 8</td>
<td>33 4 4</td>
</tr>
<tr>
<td></td>
<td>1 4 1</td>
<td>1 0 0</td>
<td>1 0 2</td>
<td>0 3 0</td>
</tr>
<tr>
<td></td>
<td>1 0 7</td>
<td>2 2 7</td>
<td>1 1 6</td>
<td></td>
</tr>
<tr>
<td>OA</td>
<td>0.83</td>
<td>0.73</td>
<td>0.66</td>
<td>0.81</td>
</tr>
<tr>
<td>Kappa</td>
<td>0.64</td>
<td>0.41</td>
<td>0.09</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Test of significance of observed differences

$$Z = \frac{|\hat{K}_m - \hat{K}_n|}{\sqrt{\text{var} (\hat{K}_m) + \text{var} (\hat{K}_n)}}.$$
Introduction

High Dimensional Kullback-Leibler Divergence

Experimental results

Conclusion
Conclusion and perspectives

Conclusion

- **Gaussian modelling** seems to be efficient to model grassland pixels distribution.
- The proposed **HDKLD is efficient** for grassland classification.
- Results are not **significantly different**.

Perspectives

- The method will be tested on a **larger dataset**.
- The method will be further extended to **multispectral data**.
- The method will be used for **unsupervised classification**.
Thank you for your attention.

Questions?
Optimal parameters have been optimized during cross-validation given this search grid:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>p-SVM</th>
<th>μ-SVM</th>
<th>KLD-SVM</th>
<th>HDKLD-SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ</td>
<td>(2^{-5}, 2^{-4}, \ldots, 2^5)</td>
<td></td>
<td>(2^8, 2^9, \ldots, 2^{12})</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td>(1, 10, 100)</td>
<td></td>
</tr>
<tr>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td>(0.80, 0.85, 0.90, 0.95, 0.99)</td>
</tr>
</tbody>
</table>